AD-A150 509 BOOTSTRAPPING THE KALMAN FILTER(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS D S STOFFER DEC 84 TR-84-51 AFOSR-TR-85-0014 F49620-82-K-0001 1/1

UNCLASSIFIED F/G 12/1 NL

END

BOOTSTRAPPING THE KALMAN FILTER

D. S. Stoffer*

Center for Multivariate Analysis

University of Pittsburgh

BOOTSTRAPPING THE KALMAN FILTER

D. S. Stoffer*

December 1984

Technical Report No. 84-51

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

*Part of the work of this author was supported by the Air Force Office of Scientific Research under Contract F49620-82-K-0001. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Bootstrapping the Kalman Filter

D.S. Stoffer*

Department of Mathematics and Statistics
University of Pittsburgh

Abstract

The bootstrap is proposed as a method for estimating the precision of forecasts and estimates of parameters of the Kalman Filter model. It is shown that when the system and the filter is in steady state the bootstrap applied to the Gaussian innovations yields asymptotically consistent standard errors. That the bootstrap works well with moderate sample sizes and supplies *robustness against departures* from normality is substantiated by emperical evidence.

Keywords: Bootstrap, Kalman Filter, Forecasting, Robustness.

*This research was supported in part by the Air Force Office of Scientific Research under Contract F49620-82-K-0001.

1. Introduction

The Kalman Filter (KF) has become an important and powerful tool for the statistician. Recently, many authors have exploited the state-space model and KF recursions for estimation and prediction of time series. For example, Jones (1980) and Harvey and Pierse (1984) use the KF to obtain maximum likelihood estimates of the parameters of ARMA processes when observations are missing. It has been suggested by Morrison and Pike (1977) and others (cf. Kendall (1973)) that the KF model provides an appropriate setting within which to parametrize smoothing and forecasting problems.

To be specific, we suppose that a $p\times 1$ vector time series $\{y_t;\ t=0,\pm 1,\pm 2,\ldots\}$ is being generated by the following dynamic system

$$y_t = x_t + v_t \tag{1.1}$$

where x_t is a zero mean, $p\times 1$ vector stationary stochastic signal, and v_t is $p\times 1$ Gaussian white noise, $v_t \sim N(0,R)$. The dynamics of the stationary signal is given by

$$x_t = \Phi x_{t-1} + w_t \tag{1.2}$$

where Φ is the $p\times p$ transition matrix and w_t is $p\times 1$ Gaussian white noise, $w_t \sim N(0,Q)$. Furthermore, $\{v_t\}$ and $\{w_t\}$ are mutually independent and we assume that the system and the filter have reached steady state. We remark that the superficially more general model in which (1.1) is replaced by

$$y_t = Mx_t + v_t$$

where M is a nonsingular known design matrix may be reduced to (1.1) by an appropriate change of bases.

Given the parameters of the model, namely, Φ, Q and R, one may obtain the minimum mean square error filter and forecasts for the system via the KF recursions. However the parameters are rarely known and hence must be estimated.

It is clear that an inexact filter model will degrade the filter performance. In fact, such an inexact model may cause the filter to diverge (cf. Jazwinski (1970), pp. 244-251). Hence, the precision of the parameter estimates must be evaluated. We propose the bootstrap as a method to evaluate the precision of parameter estimates, in particular, to handle heteroscedasticity, to provide robustness against departure from normality in the Gaussian state and observation errors, and to assist in estimating forecast errors.

Computationally simple estimates of the parameters of the KF model have been given by Anderson, Kleindorfer, Kleindorfer, and Woodroofe (1969) which we henceforth denote by AKKW. Their estimates, which are discussed in Section 4, are based on standard ergodic theory and yield strongly consistent estimates of Φ, Q and R under minor restrictions. Hence these estimates, which do not assume a Gaussian likelihood, have none of the drawbacks of the iterative maximum likelihood techniques such as Newton-Raphson or scoring (cf. Gupta and Mehra (1974)) or the EM algorithm (cf. Shumway and Stoffer (1981) and Wu (1983)) which may not converge or converge to the wrong stationary point.

We suggest that for the KF model, the AKKW estimates are the most reasonable to bootstrap since the theory driving the estimators is sound. We make bootstrapping these estimates appealing by showing in Section 5, that the bootstrap gives the right answers with large samples, that is, it is at least as sound as the conventional asymptotics. Moreover, if the investigator of such a system would rather rely on a maximum likelihood iterative scheme for parameter estimation, the AKKW estimates could be used to initialize such iterative procedures since, as will be seen, the bootstrap will work for the initial estimates.

Finally, in Section 6, we give emperical evidence of the bootstrap's importance in Kalman filtering by comparing the bootstrap to the conventional asymptotics in the cases when the likelihood is Gaussian and when the likelihood is contaminated Gaussian. An example from hydrology is also given.

Our goal is to estimate the precision of the parameter estimates of Φ, Q and R, as well as the precision of the forecasts $\hat{x}_{n+1}, \hat{x}_{n+2}, \ldots, \hat{x}_{n+k}$. The techniques used here are based on the bootstrap (cf. Efron (1979)) and the methods used in bootstrapping least squares estimates discussed in Bickel and Freedman (1981), Freedman (1981) and Freedman and Peters (1984). It is noted in the above references that in regression models (static or dynamic), it is appropriate to resample the centered residuals after estimating the parameters. This is not possible in the present model (1.1) and (1.2) since the signal is not observable. However, we may base the procedure on the innovations which are obtained by taking the conditional expectation of the signal given the data. Hence, the bootstrap procedure will involve the resampling of the innovation sequence

$$r_t^{t-1} = y_t - x_t^{t-1}, \qquad t = 1, \ldots, n \tag{1.3}$$

where by x_t^{t-1} we mean $E(x_t | y_1, \ldots, y_{t-1})$. Of course x_t^{t-1} will be obtained recursively via the KF.

Under the conditions stated in the next section we will be able to put this problem into the nonlinear regression context as discussed in Efron (1979, Section 7). That is, we may write

$$y_t = g_t(\Phi,Q,R,x_t,y_1,\ldots,y_{t-1}) + \varepsilon_t, \qquad t=1,2,\ldots,n$$

where ε_t are iid zero mean random vectors (namely, the innovations) and $g_t(\cdot)$ is a particularly complicated, but known, nonlinear function of the parameters Φ, Q, and R, the signal x_t, and the data $y_1,\ldots,y_{t-1}$. In particular, $g_t(\cdot) = x_t^{t-1}$, the filtered value of the signal.

In the next section we give conditions under which we are able to bootstrap the innovations, (1.3). The bootstrap procedure is given in Section 3.

2. The Steady-State Innovation Sequence

Throughout the remainder of this paper we make the following assumptions on the p×p parameter matrices: (A1) Q and R are positive definite, and (A2) Φ is nonsingular with spectal norm, $\rho(\Phi)$, less than unity. These conditions ensure the asymptotic global stability of the KF (cf. Deyst and Price (1968)).

The steady-state KF recursions are given by (cf. Jazwinski (1970))

$$K = P(P+R)^{-1}, \tag{2.1a}$$

$$P = \Phi[P-P(P+R)^{-1}P]\Phi' + Q, \tag{2.1b}$$

$$x_{t+1}^t = \Phi x_t^t, \tag{2.1c}$$

$$x_t^t = x_t^{t-1} + K(y_t - x_t^{t-1}). \tag{2.1d}$$

In the KF above, K is the steady-state gain matrix, P is the steady-state filter error, $P = E\{(x_t-x_t^{t-1})(x_t-x_t^{t-1})'\}$, and $x_t^{t-1} = E(x_t|y_1,\ldots,y_{t-1})$ is the steady-state filter estimate of x_t based on the data $y_1,\ldots,y_{t-1}$.

<u>Proposition 2.1</u> Under steady-state and optimal filtering, the p×1 vector innovation sequence

$$r_t^{t-1} = y_t - x_t^{t-1}, \qquad t = 1,\ldots,n \tag{2.2}$$

is a zero-mean, white Gaussian sequence with covariance matrix P+R.

Proof Write $r_t^{t-1} = e_t + v_t$ where $e_t = x_t - x_t^{t-1}$ and note that $E(e_t) = E(v_t) = 0$. The r_t^{t-1} are Gaussian since they are linear combinations of Gaussian random vectors. To establish the orthogonality of the innovations, it is easy to see that while $r_t^{t-1}(= y_t - x_t^{t-1})$ is in the linear space spanned by $\{y_1,\ldots,y_t\}$, $r_t^{t-1}(= e_t + v_t)$ is orthogonal to the linear space spanned by $\{y_1,\ldots,y_{t-1}\}$. Hence, for $s < t$,

$$E(r_s^{s-1} r_t^{t-1'}) = E\{r_s^{s-1} E(r_t^{t-1'} | y_1,\ldots,y_s \quad)\} = 0.$$

Also, since e_t and v_t are uncorrelated we have that

$$\mathrm{Cov}(r_t^{t-1}) = \mathrm{Cov}(e_t) + \mathrm{Cov}(v_t) = P+R. \qquad \square$$

As a final remark, we note that via (2.1c) and (2.1d), we may write y_t in terms of the steady-state innovations as

$$y_t = \sum_{j=1}^{\infty} \Phi^j K \, r_{t-j}^{t-j-1} + r_t^{t-1} \tag{2.3}$$

which follows from the fact that $\|\Phi^j\| \to 0$ exponentially fast as $j \to \infty$ since $\rho(\Phi) < 1$, where $\|\Phi\|^2 = \mathrm{trace}(\Phi'\Phi)$. This result will be useful in establishing the bootstrap procedure.

3. The Bootstrap Estimate of Precision

As previously mentioned, the bootstrap technique will be employed by resampling the steady-state innovation sequence. Recall that under optimal filtering the innovation sequence r_t^{t-1}, $t = 1,\ldots,n$ is $p\times 1$ Gaussian white noise,

$r_t^{t-1} \sim N(0,P+R)$ where P is the steady-state error covariance matrix given in (2.1b).

The bootstrap procedure begins be estimating the parameters $\Theta = \{\Phi,Q,R\}$ of the model (1.1), (1.2) by some optimal procedure as mentioned in the Introduction. We shall discuss a particular method in Section 4. Call these estimates $\hat{\Theta} = \{\hat{\Phi},\hat{Q},\hat{R}\}$.

From these preliminary estimates obtain a suboptimal innovation sequence by filtering (cf. 2.1) under $\hat{\Theta}$. Call this innovation sequence $\hat{r}_t^{t-1}$. Make the sequence $\{\hat{r}_t^{t-1}\}_{t=1}^n$ independent and identically distributed with distribution equal to the emperical distribution by putting mass n^{-1} on each innovation $\hat{r}_t^{t-1}$, $t=1,\ldots,n$.

Next, draw a "bootstrap sample" of innovations, r_t^{*t-1}, $t=1,\ldots,n$ by independent random sampling of the residuals $\hat{r}_t^{t-1}$. That is, sample the $\hat{r}_t^{t-1}$,n times, with replacement from $\{\hat{r}_1^0,\hat{r}_2^1,\ldots,\hat{r}_n^{n-1}\}$. From this we obtain a "bootstrap sample" of data $y_1^*,\ldots,y_n^*$ by setting (cf. 2.3)

$$y_t^* = r_t^{*t-1} + \sum_{j=1}^{t-1} \hat{\Phi}^j \hat{K} \hat{r}_{t-j}^{*t-j-1}, \qquad t=1,\ldots,n \tag{3.1}$$

where $\hat{K}$ is the estimated gain matrix obtained via filtering under parameters $\hat{\Theta}$.

We make the following suggestions before proceeding with step (3.1). First, as suggested in Freedman (1981), one should center the residuals $\hat{r}_t^{t-1}$ before resampling them so that the emperical distribution puts mass n^{-1} on $\hat{r}_t^{t-1} - \hat{\mu}_n$ where $\hat{\mu}_n = n^{-1}\sum_{t=1}^n \hat{r}_t^{t-1}$. Second, we suggest checking whether the innovations are white. It is known that a suboptimal filter produces correlated innovations (see, for example, Mehra (1970)) and hence this is a check on the "goodness" of the estimates. Various methods are available for testing the whiteness of the innovations many of which are listed in Mehra (1970).

Now, suppose that the bootstrap data $\{y_1^*,\ldots,y_n^*\}$ come from the model

$$y_t^* = x_t + v_t^*, \qquad t \geq 1, \tag{3.2a}$$

$$x_t = \Phi^* x_{t-1} + w_t^*, \quad t \geq 1, \tag{3.2b}$$

where v_t^* is $p\times 1$ Gaussian white noise $v_t^* \sim N(0,R^*)$ and is independent of w_t^* which is $p\times 1$ Gaussian white noise $w_t^* \sim N(0,Q^*)$. Assume the parameters $\Theta^* = \{\Phi^*,Q^*,R^*\}$ are unknown and to be estimated.

The parameters Θ^* are then estimated by the initial optimal procedure to produce estimates $\hat{\Theta}^* = \{\hat{\Phi}^*,\hat{Q}^*,\hat{R}^*\}$. Then, the suboptimal innovation sequence is resampled and the bootstrap procedure is reiterated.

The entire process is repeated some large number "L" of times obtaining L bootstrap replications $\hat{\Theta}_1^*,\hat{\Theta}_2^*,\ldots,\hat{\Theta}_L^*$. The distribution of the errors

$$\hat{\Phi}^* - \hat{\Phi}, \; \hat{Q}^* - \hat{Q}, \; \hat{R}^* - \hat{R} \tag{3.3}$$

are then computed to give an approximation as to the distribution of

$$\hat{\Phi} - \Phi, \; \hat{Q} - Q, \; \hat{R} - R. \tag{3.4}$$

The bootstrap distribution of the errors (3.3) may then be used to obtain confidence regions and tests of hypotheses about the parameters Θ. Justification of this procedure is given in Section 5, Theorem 5.1.

Forecasting k steps into the future, say $x_{n+j}^n = E(x_{n+j}|y_1,\ldots,y_n)$, $j=1,2,\ldots,k$ is easily accomplished via the filter equations (2.1), namely

$$x_{n+j}^{n} = \Phi^{j} x_{n}^{n}, \qquad j = 1,\ldots,k. \tag{3.5}$$

The suboptimal forecasts will be obtained via the KF under parameter estimates $\hat{\Theta}$ so that

$$\hat{x}_{n+j}^{n} = \hat{\Phi}^{j} \hat{x}_{n}^{n}, \qquad j = 1,\ldots,k \tag{3.6}$$

will be the actual forecasts. If at each bootstrap replication we obtain

$$\{\hat{x}_{n+1}^{n*},\ldots,\hat{x}_{n+k}^{n*}\}^{(1)},\ldots,\{\hat{x}_{n+1}^{n*},\ldots,\hat{x}_{n+k}^{n*}\}^{(L)} \tag{3.7}$$

we may extract the emperical distribution of the forecast residuals

$$\hat{x}_{n+j}^{n*} - \hat{x}_{n+j}^{n}, \qquad j = 1,\ldots,k \tag{3.8}$$

which can then be used to approximate the distribution of the actual forecast errors

$$\hat{x}_{n+j}^{n} - x_{n+j}^{n}, \qquad j = 1,\ldots,k. \tag{3.9}$$

From the distributions of (3.8) we may obtain prediction regions for the forecasts (3.5).

4. Consistent Parameter Estimates

In this section we give the details of the consistent estimation of the parameters of the KF model (1.1), (1.2). Recall that the system is in steady-state and that the parameters $\Theta = \{\Phi, Q, R\}$ satisfy the conditions (A1) and (A2) given in Section 2.

The observation sequence may then be written as

$$y_t = v_t + \sum_{j=0}^{\infty} \Phi^{j} w_{t-j} \tag{4.1}$$

from which it follows that y_t is a strictly stationary, zero-mean Gaussian sequence with covariance matrix $R + \sum_{j=0}^{\infty}\Phi^j Q\Phi^{j'}$.

The estimates of Φ, Q and R are based on the matrices

$$C_n(h) = n^{-1}\sum_{t=h+1}^{n} y_t y'_{t-h}, \qquad h \geq 0 \tag{4.2}$$

which in view of (4.1) and the Ergodic Theorem are ergodic and converge almost surely (a.s.) to

$$\Gamma(h) = \Phi^h \sum_{j=0}^{\infty}\Phi^j Q\Phi^{j'} + \delta_{h,0}R, \qquad h \geq 0 \tag{4.3}$$

where $\delta_{h,0}$ is the Kronecker δ and $\Gamma(h) = \mathrm{Cov}(y_t, y_{t-h})$.

An estimate of Φ is suggested by the fact that $E(y_t y'_{t-2}) = \Phi E(y_{t-1}y'_{t-2})$, $t \geq 3$, namely,

$$\hat{\Phi}_n = \left(\sum_{t=3}^{n} y_t y'_{t-2}\right)\left(\sum_{t=3}^{n} y_{t-1} y'_{t-2}\right)^+, \qquad n \geq 3 \tag{4.4}$$

where by $+$ we mean generalized inverse. Estimates of Q and R are suggested by the facts that

$$B(1) = E\{(y_t - \Phi y_{t-1})(y_t - \Phi y_{t-1})'\}$$
$$= Q + R + \Phi R\Phi'$$
$$B(2) = E\{(y_t - \Phi^2 y_{t-2})(y_t - \Phi^2 y_{t-2})'\}$$
$$= Q + R + \Phi Q\Phi' + \Phi^2 R\Phi^{2'}$$

which yield

$$R = \tfrac{1}{2}\{B(1) + \Phi^{-1}(B(1) - B(2))\Phi^{-1'}\} \tag{4.5}$$

$$Q = B(1) - R - \Phi R\Phi' \tag{4.6}$$

provided that Φ is nonsingular.

We now state the following theorems which follow directly from the almost sure convergence of (4.2) to (4.3). Denote the spectral norm of Φ by $\rho(\Phi)$.

Theorem 4.1 If $\rho(\Phi) < 1$, and if Φ and Q are nonsingular, then $\hat{\Phi}_n$ given in (4.4) is strongly consistent for Φ.

Theorem 4.2 If $\hat{\Phi}_n$ is strongly consistent for Φ and if $\rho(\Phi) < 1$ then

$$\hat{B}_n(i) = n^{-1}\sum_{t=3}^{n}(y_t - \hat{\Phi}_n^i y_{t-i})(y_t - \hat{\Phi}_n^i y_{t-i})' \quad n \geq 3 \tag{4.7}$$

is strongly consistent for B(i), $i = 1,2$. Hence (4.4) and (4.7) provide strongly consistent estimates for R and Q via (4.5) and (4.6) provided that Φ and Q are nonsingular.

Next, we exhibit the behavior of the suboptimal filter and forecasts in the following theorems and corollary (cf. Theorem 2.4, Theorem 2.5, and Corollary 2.4 of AKKW). Denote positive (semi)-definite by p.(s).d.

Theorem 4.3 If Q is p.d. and if $\hat{\Phi}_n, \hat{Q}_n, \hat{R}_n$ are strongly consistent estimates of Φ,Q,R respectively for which $\hat{Q}_n$ is p.d. and R_n is p.s.d. for all $n \geq 1$, then $\hat{P}_n \to P$ and $\hat{K}_n \to K$ a.s. as $n \to \infty$ where $\hat{P}_n$ and $\hat{K}_n$ are the estimates of the steady-state filter covariance and gain matrices, respectively.

Theorem 4.4 Let the hypotheses of Theorem 4.3 be satisfied. If in addition, $\rho(\Phi) < 1$, then

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}|x_t^t - \hat{x}_t^t|^2 = 0 \qquad \text{a.s.}$$

Corollary 4.1 If the hypotheses of Theorem 4.4 are satisfied, then

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}|x_{t+k}^{t} - \hat{\Phi}_{n}^{k}\hat{x}_{t}^{t}|^2 = 0 \quad \text{a.s.}$$

for any $k \geq 1$.

Theorem 4.4 and its corollary will be useful in establishing the bootstrap principle which is discussed in the next section. We conclude this section with a statement about the limiting law of the matrices $C_n(h)$ given by (4.2). The following theorem follows from Hannan (1970, p. 228).

Theorem 4.5 Let y_t be generated by (4.1). Let $c_{ij}(h)$ and $\gamma_{ij}(h)$ denote the ijth element of $C_n(h)$ defined in (4.2) and $\Gamma(h)$ defined in (4.3), respectively. Then, for any integer $H > 0$, the joint law of

$$\sqrt{n}\,\{c_{ij}(h) - \gamma_{ij}(h)\} \qquad i,j = 1,\ldots,p;\ h = 1,2,\ldots,H$$

converges to that of a zero-mean normal, the asymptotic covariance between $c_{ij}(m)$ and $c_{k\ell}(n)$ being

$$\sum_{r=-\infty}^{\infty}\{\gamma_{ik}(r)\gamma_{j\ell}(r+n-m) + \gamma_{i\ell}(r+n)\gamma_{jk}(r-m)\}. \tag{4.8}$$

In general, the covariance of the $c_{ij}(h)$ involves the forth order cumulants of the process which vanish in this case since y_t is Gaussian. For more details concerning the asymptotic covariances of $c_{ij}(h)$, the reader is referred to Hannan (1970) or Anderson (1971).

5. The Bootstrap Principle

In this section we justify the techniques established in Section 3. The mathematical tools used in this section are those of Bickel and Freedman (1981).

The estimate of Φ given in (4.4) may be rewritten as

$$\hat{\Phi}_n = \Phi + \Delta_n$$

where

$$\Delta_n = [n^{-1}\sum_{t=3}^{n}(y_t - \Phi y_{t-1})y'_{t-2}][n^{-1}\sum_{t=3}^{n} y_{t-1}y'_{t-2}]^{+}. \tag{5.1}$$

It is clear that the asymptotic distribution of $\sqrt{n}(\hat{\Phi}_n - \Phi) = \sqrt{n}\,\Delta_n$ may then be established via Theorem 4.5 and the almost sure convergence of (4.2) to (4.3). Similarly, in view of (4.5) and (4.6), the asymptotic distributions of $\sqrt{n}(\hat{R}_n - R)$ and $\sqrt{n}(\hat{Q}_n - Q)$ may be established from the asymptotic distributions of $\sqrt{n}(\hat{B}_n(i) - B(i))$, $i = 1,2$, which in turn may be established via Theorem 4.5 and the almost sure convergence of (4.2) to (4.3).

For example, in the univariate case (p=1) it is seen that

$$n\ \mathrm{Var}(\hat{\phi}_n - \phi) \to \{V(2,2) + \phi^2 V(1,1) - 2\phi\ V(1,2)\}\gamma^{-2}(1) \tag{5.2a}$$

$$n\ \mathrm{Var}(\hat{B}_n(1) - B(1)) \to \{(1+\phi^2)^2 V(0,0) + 4\phi^2 V(1,1) - 4\phi(1+\phi^2)V(0,1)\} \tag{5.2b}$$

$$n\ \mathrm{Var}(\hat{B}_n(2) - B(2)) \to \{(1+\phi^4)^2 V(0,0) + 4\phi^4 V(2,2) - 4\phi^2(1+\phi^4)V(0,2)\} \tag{5.2c}$$

where we have the set

$$V(i,j) = \sum_{r=-\infty}^{\infty}\{\gamma(r)\gamma(r+i-j) + \gamma(r+j)\gamma(r-i)\}.$$

Let $\Sigma(1)$ be (5.2b) and $\Sigma(2)$ be (5.2c). It can then be shown that

$$n\ \mathrm{Var}(\hat{R}_n - R) \to \{(1+\phi^{-2})^2\Sigma(1) + \phi^{-4}\Sigma(2) - 2(1+\phi^{-2})\phi^{-2}\Sigma(12)\}\frac{1}{4} \tag{5.2d}$$

$$n\ \mathrm{Var}(\hat{Q}_n - Q) \to \{\Sigma(1) + (1+\phi^2)^2\Sigma(R) - 2(1+\phi^2)\Sigma(R1)\} \tag{5.2e}$$

where we have set $\Sigma(R)$ to be (5.2d), $n\ \mathrm{Cov}(\hat{B}_n(1), \hat{B}_n(2)) \to \Sigma(12)$ where

$$\Sigma(12) = \{(1+\phi^2)(1+\phi^4)V(0,0) - 2\phi^2(1+\phi^2)V(0,2) - 2\phi(1+\phi^4)V(1,0) + 4\phi^3 V(1,2)\},$$

and $n\ \mathrm{Cov}(\hat{R}_n, \hat{B}_n(1)) \to \Sigma(R1)$ where

$$\Sigma(R1) = \frac{1}{2}(1+\phi^{-2})\Sigma(1) - \frac{1}{2}\phi^{-2}\Sigma(12).$$

Now, let starred variables denote those obtained via the bootstrap sample $\{y_1^*,\ldots,y_n^*\}$. In this manner we denote

$$C_n^*(h) = n^{-1} \sum_{t=h+1}^{n} y_t^* y_{t-h}^{*\prime}, \qquad h \geq 0 \tag{5.3}$$

as the bootstrap counterpart of $C_n(h)$, equation (4.2), upon which all the estimates $\hat{\Phi}_n$, $\hat{Q}_n$, $\hat{R}_n$, $\hat{B}_n(1)$, and $\hat{B}_n(2)$ are based.

The bootstrap principle given in Section 3 is now stated in the following theorem.

Theorem 5.1 Along almost all sample sequences, as $n\to\infty$, conditionally on the data, for all $h \geq 0$,

(1) $C_n^*(h) \to \Gamma(h)$ in conditional probability, and

(2) the conditional law of $\sqrt{n}(C_n^*(h)-C_n(h))$ merges with the unconditional law of $\sqrt{n}(C_n(h)-\Gamma(h))$.

From Theorem 5.1 we may then establish that the conditional laws of $\sqrt{n}\ \Delta_n^*$, $\sqrt{n}(\hat{B}_n^*(i)-\hat{B}_n(i))$ $i=1,2$ merge with the unconditional law of $\sqrt{n}\ \Delta_n$ and $\sqrt{n}(\hat{B}_n(i)-B(i))$ $i=1,2$ by what is essentially Slutsky's theorem.

As previously suggested, the proof of Theorem 5.1 is based on the Mallows metrics argument of Bickel and Freedman (1981) and is fashioned after Freedman (1984). If R^p is a p-dimensional space equipped with the Euclidean norm $|\cdot|$ and $\alpha \geq 1$, then $d_\alpha^p(\mu,\nu)$ is the distance between probability measures μ and ν in R^p defined as the infimum of $\acute{E}\{|U-V|^\alpha\}^{1/\alpha}$ over all pairs of random vectors U with law μ and V with law ν.

Before proving Theorem 5.1 we establish the following lemmas.

Lemma 5.1 Let $\theta=(\Phi,Q,R)$ and $\hat{\theta}_n=(\hat{\Phi}_n,\hat{Q}_n,\hat{R}_n)$ satisfy the conditions of Theorem 4.4. Let $\hat{F}_n$ be the emperical distribution function (e.d.f.) of the suboptimal innovations $\hat{r}_t^{t-1}$, $t=1,\dots,n$ generated by $\hat{\theta}_n$ and let F_n be the e.d.f. of the optimal innovations r_t^{t-1} $t=1,\dots,n$ generated by θ. Then $d_2^p(\hat{F}_n,F_n)\to 0$ almost surely (a.s.) as $n\to\infty$.

Proof Noting that $r_t^{t-1}=y_t-x_t^{t-1}$ and $\hat{r}_t^{t-1}=y_t-\hat{x}_t^{t-1}$, in view of Theorem 4.4, we have

$$d_2^p(\hat{F}_n,F_n)^2\le n^{-1}\sum_{t=1}^n|\hat{r}_t^{t-1}-r_t^{t-1}|^2$$

$$=n^{-1}\sum_{t=1}^n|x_t^{t-1}-\hat{x}_t^{t-1}|^2\to 0 \quad \text{a.s.}$$

as $n\to\infty$. □

Lemma 5.2 Let F_n be the e.d.f. of the optimal innovations, r_t^{t-1}, $t=1,\dots,n$ and let F be the common distribution of r_t^{t-1}. Then $d_2^p(F_n,F)\to 0$ a.s. as $n\to\infty$.

Proof Since the optimal steady-state innovations are iid (cf. Proposition 2.1), this follows from Lemma 8.4 of Bickel and Freedman (1981). □

Now, let $\psi_{n,h}(F)$ be the law of $C_n(h)$, $h\ge 0$, when the law of r_t^{t-1} is F. Metrize the ψ's by $d_1^{p\times p}$ and the F's by d_2^p. For notational convenience, we drop the subscript h from the ψ's and drop the superscript $t-1$ from the r_t's.

Lemma 5.3 The $\psi_n(F)$ are equiuniformly continuous functions of F on

$$S=\{F:\int_{R^p}|r|^2dF(r)\le c^2<\infty\}. \tag{5.4}$$

Proof The proof of this lemma is similar to Freedman (1984, Lemma 6.3). Fix F and F^* in S. Construct iid random vectors (r_t,r_t^*), $t=1,\dots,n$ so that r_t has law F and r_t^* has law F^*, and

$$d_2^p(F,F^*)^2=E\{|r_t-r_t^*|^2\},$$

See Bickel and Freedman (1981, Lemma 8.1). Build y_t from the r_t and y_t^* from the r_t^* as in (2.3). Then, for $h \geq 0$

$$d_1^{p\times p}[\psi_n(F),\psi_n(F^*)] \leq E\{|n^{-1}\sum_{t=h+1}^{n}(y_t y_{t-h}' - y_t^* y_{t-h}^{*\prime})|\}$$

$$\leq E\{|y_t y_{t-h}' - y_t^* y_{t-h}^{*\prime}|\}$$

$$\leq E\{|y_t|\cdot|y_{t-h}-y_{t-h}^*|\} + E\{|y_t-y_t^*|\cdot|y_{t-h}^*|\}. \qquad (5.5)$$

Now, by the Cauchy-Schwartz inequality and the remarks following equation (4.1)

$$E\{|y_t|\cdot|y_s-y_s^*|\}^2 \leq E\{|y_t|^2\}\, E\{|y_s-y_s^*|^2\}$$

$$\leq c^2\, E\{|y_s-y_s^*|^2\}.$$

Using the fact that if U_j are independent random vectors, then

$$E\{|\Sigma_j U_j|^2\} \leq \Sigma_j E\{|U_j|^2\} + |\Sigma_j\, E\{U_j\}|^2$$

we have that in view of equation (2.3)

$$E\{|y_s-y_s^*|^2\} = E\{|\sum_{j=1}^{\infty} \Phi^j K(r_{s-j}-r_{s-j}^*) + (r_s-r_s^*)|^2\}$$

$$\leq E\{|\delta\ |^2\} + E\{|\delta\ |^2\}\, \frac{\rho k}{1-\rho^2} + |E\{\delta\ \}|^2 \left[\left(\frac{\rho k}{1-\rho}\right) + 1\right]^2 \qquad (5.6)$$

where

$$\delta = r_1 - r_1^*, \quad \rho = \|\Phi\| < 1, \quad \text{and} \quad k = \|K\|.$$

It is clear that (5.6) is small if F and F^* are close in d_2. □

Now let $\Omega_{n,h}(F)$ be the law of $\sqrt{n}\, C_n(h)$, $h \geq 0$, when the law of r_t is F.

Metrize the Ω's by $d_2^{p\times p}$ and the F's by d_2^p. Again, we drop subscript h from the Ω's.

Lemma 5.4 The $\Omega_n(F)$ are equiuniformly continuous functions of F on S (cf. 5.4). The proof of this lemma follows by bounding $d_2^{p\times p}[\Omega_n(F),\Omega_n(F^*)]$ in much the same way as $d_1^{p\times p}[\psi_n(F), \psi_n(F^*)]$ was bounded in Lemma 5.3.

For example, put $A_t(h) = y_t y'_{t-h} - y^*_t y^{*'}_{t-h}$, then

$$d_2^{p\times p}[\Omega_n(F),\ \Omega_n(F^*)]^2 \leq E\{|\sqrt{n}^{-1}\textstyle\sum_{t=h+1}^{n} A_t(h)|^2\}$$

$$\leq E\{|A_t(h)|^2\} + n^{-1} \textstyle\sum_{t\neq s} E\{|A_t(h)\cdot A_s(h)|\}. \qquad (5.7)$$

The first part of (5.7) is bounded as in (5.5) to (5.6). Furthermore, by the Cauchy-Schwartz inequality

$$E\{|A_t(h)\cdot A_s(h)|\}^2 \leq E\{|A_t(h)|^2\}^2$$

which may be bounded (independent of t) as in (5.6).

The proof of part (1) of Theorem 5.1 now follows from Lemmas 5.1, 5.2, and 5.3. The proof of part (2) of Theorem 5.1 similarly follows from Lemmas 5.1, 5.2, and 5.4. For example, part (1) follows since the conditional law of $C_n^*(h)$ given the data differs little in the sense of $d_1^{p\times p}$ from the unconditional law of $C_n(h)$ by Lemma 5.3, because the e.d.f. of the suboptimal innovations, $\hat{F}_n$, differs little in the sense of d_2^p from the law of the optimal innovations, F, by the combination of Lemmas 5.1 and 5.2. This concludes our proof and this section.

6. Examples

In this section we submit three univariate examples as emperical evidence of the bootstrap's performance as an aid in Kalman filtering.

6.1 Simulation: Normal Errors

In this example we generated $n = 100$ Gaussian observations from the model (1.1), (1.2) with $\phi = 0.8$, $Q = 4.0$, and $R = 1.0$. Each parameter estimate was bootstrapped $L = 100$ times. Table 6.1.1 gives the summary results of 30 runs; the true asymptotic values are obtained via (5.2) and are listed as standard errors. The sample standard errors obtained by inserting estimates for the true values into (5.2) are also listed.

TABLE 6.1.1

	$SE(\hat{\phi})$		$SE(\hat{Q})$		$SE(\hat{R})$	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Bootstrap	0.106	0.030	1.242	0.256	0.779	0.361
Sample	0.153	0.068	1.749	0.421	1.070	0.198
True	0.083		1.269		0.843	

Table 6.1.1: Summary of the standard errors of 30 runs of the bootstrap procedure ($L=100$) with $n = 100$ normal observations from the model (1.1), (1.2) with $\phi = 0.8$, $Q = 4.0$, $R = 1.0$.

Table 6.1.2 contains the summary of the centered innovations from the 30 runs described above. The standard deviation of the 100 innovations as well as the first five autocorrelations are listed. From (2.1) we may calculate the true standard deviation of the innovations, $\sqrt{P+R} = 2.35$. Also, for large sample sizes n from iid data, the autocorrelations have approximate means and standard deviations of $-1/n$ and $1/\sqrt{n}$, respectively.

TABLE 6.1.2

		Sample	True
STANDARD DEVIATION		2.380 ± 0.168	2.350
AUTOCORRELATION			
LAG:	1	-0.006 ± 0.213	-0.01 ± 0.10
	2	-0.007 ± 0.069	-0.01 ± 0.10
	3	0.033 ± 0.098	-0.01 ± 0.10
	4	-0.025 ± 0.102	-0.01 ± 0.10
	5	-0.016 ± 0.121	-0.01 ± 0.10

Table 6.1.2: Summary (mean $\pm$ st.dev.) of the innovations from 30 runs of the Kalman Filter for $n = 100$ normal observations from the model (1.1), (1.2) with $\phi = 0.8$, $Q = 4.0$, $R = 1.0$.

6.2 Simulation: Contaminated Normal Errors

In this example we generated $n = 100$ contaminated normal observations from the model (1.1), (1.2) with $\phi = 0.8$, $Q = 4.0(90\%) + 16.0(10\%) = 5.2$, and $R = 1.0(90\%) + 9.0(10\%) = 1.8$. That is, the state errors are $90\%N(0,4) + 10\%N(0,16)$ and the observation errors are $90\%N(0,1) + 10\%N(0,9)$. The true asymptotic variances of the estimates must now include the forth cumulants of the error sequences (cf. Hannan (1970) or Anderson (1971)) so that $V(i,j)$ of (5.2) is adjusted to $V(i,j) + \kappa_w \gamma_x(i)\gamma_x(j)\sigma_w^{-4} + \kappa_v \sigma_v^{-2}\delta_{i,0}\delta_{j,0}$ where κ_w and κ_v are the forth cumulants of the state and observation error sequences, respectively, $\gamma_x(i)$ is $\mathrm{Cov}(x_t, x_{t+i})$, σ_w^2 and σ_v^2 are the variances Q and R of the state and observation error sequences and $\delta_{i,0}, \delta_{j,0}$ are the Kronecker deltas.

Table 6.2.1 gives the summary results of 30 runs in which each parameter estimate was bootstrapped $L = 100$ times. The true asymptotic standard errors are obtained as described above and the sample standard errors are calculated by inserting estimates for the true values in (5.2).

TABLE 6.2.1

	SE($\hat{\phi}$)		SE($\hat{Q}$)		SE($\hat{R}$)	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
Bootstrap	0.110	0.024	2.141	0.442	1.383	0.437
Sample	0.174	0.083	2.793	0.674	1.758	0.374
True	0.086		1.848		1.268	

Table 6.2.1: Summary of the standard errors of 30 runs of the bootstrap procedure ($L = 100$) with $n = 100$ observations from the model (1.1), (1.2) with $\phi = 0.8$, $Q = 5.2$, $R = 1.8$ and 10% contaminated errors.

The summary of the centered innovations obtained in this simulation is given in Table 6.2.2. In this case the true standard deviation of the innovations is $\sqrt{P+R} = 2.74$.

TABLE 6.2.2

		Sample	True
STANDARD DEVIATION		2.754 ± 0.234	2.737
AUTOCORRELATION			
LAG:	1	-0.014 ± 0.019	-0.01 ± 0.10
	2	-0.009 ± 0.061	-0.01 ± 0.10
	3	0.032 ± 0.102	-0.01 ± 0.10
	4	-0.018 ± 0.095	-0.01 ± 0.10
	5	-0.026 ± 0.719	-0.01 ± 0.10

Table 6.2.2: Summary (mean $\pm$ st.dev.) of the innovations from 30 runs of the Kalman Filter for $n = 100$ observations from the model (1.1), (1.2) with $\phi = 0.8$, $Q = 5.2$, $R = 1.8$ and 10% contaminated errors.

6.3 Annual Flows of the Göta River, Sweden

The following data are the annual flows of the Göta River near Sjötop-Vänersburg for the period 1901-1950. The data are in the form of modular values (actual annual flows divided by the mean) as given by Salas et al. (1980).

Years										
01-10	0.935	0.662	0.950	1.121	0.880	0.802	0.856	1.080	0.959	1.345
11-20	1.153	0.929	1.158	0.957	0.705	0.905	1.000	0.918	0.907	0.991
21-30	0.994	0.701	0.692	1.086	1.306	0.895	1.149	1.297	1.168	1.218
31-40	1.209	0.974	0.834	0.638	0.991	1.198	1.091	0.892	1.020	0.869
41-50	0.772	0.606	0.739	0.813	1.173	0.916	0.880	0.601	0.720	0.955

Model (1.1), (1.2) was fitted to the data and the standard errors of the estimates were obtained by conventional asymptotics (sample) and by the bootstrap with $L = 100$ iterations. The results are given in Table 6.3.1.

TABLE 6.3.1

	Parameters		
	ϕ	Q	R
1. Estimates	0.9902	0.0273	0.0067
2. Standard Errors:			
Sample	2.5410	0.0714	0.0342
Bootstrap	0.0916	0.0088	0.0051

Table 6.3.1: Summary of parameter estimates and estimated standard errors, sample and bootstrap ($L = 100$), obtained by fitting (1.1), (1.2) to the Göta River data.

We note that the standard errors obtained via conventional asymptotics are unreasonably large and hence would be useless to the investigator of this system. This situation is probably due to the fact that the signal is close to nonstationarity ($\hat{\phi} = 0.99$). The bootstrap however, provides more reasonable estimates for the precision of the parameter estimates.

The bootstrap also allows us to obtain prediction intervals. We predicted x_{51} to be $\hat{\phi}\hat{x}_{50}^{50} = 0.9029$, where $\hat{x}_{50}^{50} = 0.9118$ is obtained by filtering under $\hat{\phi}$, $\hat{Q}$, and $\hat{R}$. The filter was initialized by the mean and the variance of the data. The bootstrap estimate of the prediction error $\widehat{SE}(\hat{x}_{51})$, is thus $\widehat{SE}(\hat{\phi})\ \hat{x}_{50}^{50} = 0.0835$.

The steady state filter error $P^{1/2} = E^{1/2}(x_t - x_t^{t-1})^2$ was estimated by filtering under $\hat{\phi}$, $\hat{Q}$, and $\hat{R}$ and a value of 0.1847 was obtained. The standard error of the 50 innovations $\hat{r}_t^{t-1} = (y_t - \hat{x}_t^{t-1})$ $t = 1,\ldots,50$, was 0.202 and the estimated autocorrelation function for the first 12 lags is: -0.066, -0.374, -0.037, 0.044, -0.157, 0.010, 0.101, 0.073, -0.069, 0.033, 0.057, and -0.066.

References

[1] Anderson, T.W. (1971). The Statistical Analysis of Time Series, Wiley, New York.

[2] Anderson, W.N., Kleindorfer, G.B., and Kleindorfer, P.R., and Woodroofe, M.B. (1969). Consistent estimates of the parameters of a linear system. Ann. Math. Stat., 6, 2064-2075.

[3] Bickel, P.J. and Freedman, D.A. (1981). More on bootstrapping regression models, Technical Report, Statistical Dept., Univeristy of California, Berkeley.

[4] Deyst, J.J. and Price, C.F. (1968). Conditions ofr asymptotic stability of the discrete minimum variance linear estimator, IEEE Trans. on Aut. Control, AC-13, 702-705.

[5] Efron, B. (1979). Bootstrap methods: another look at the jackknife, Annals of Statistics, 7, 1-26.

[6] Freedman, D.A. (1981). Bootstrapping regression models, Annals of Statistics, 9, 1218-1228.

[7] Freedman, D.A. (1984). On bootstrapping two-stage least square estimates in stationary linear models, Annals of Statistics, 12, 827-842.

[8] Freedman, D.A. and Peters,.S.C. (1984). Bootstrapping a regression equation: some empirical results, Journal of the American Statistical Association, 79, 97-106.

[9] Gupta, N.K. and Mehra, R.K. (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. IEEE Trans. Aut. Cont., AC-19, 774-783.

[10] Hannan, E.J. (1970). Multiple Time Series, Wiley, New York.

[11] Harvey, A.C. and Pierse, R.G. (1984). Estimating missing observations in economic time series, Journal of the American Statistical Association, 79, 125-131.

[12] Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

[13] Jones, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics, 22, 389-395.

[14] Kendall, M.G. (1973). Time Series, Hafner Pressm New York.

[15] Mehra, R.K. (1970). On the identification of variances and adaptive Kalman filtering, IEEE Trans. on Aut. Control, AC-15, 175-184.

[16] Morrison, G.W. and Pike, D.H. (1977). Kalman filtering applied to statistical forecasting, Management Science, 23, 768-774.

[17] Salas, J.D., Delleur, J.W., Yevjevich, V. and Lane, W.L. (1980). Applied Modeling of Hydrologic Time Series, Water Resources Publications, Littleton, Colorado.

[18] Shumway, R.H. and Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. J. Time Series Anal., 3, 253-264.

[19] Wu, C.F. (1983). On the convergence of the EM algorithm. Annals of Statistics, 11, 95-103.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S): 84-51

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 85-0014

6a. NAME OF PERFORMING ORGANIZATION: University of Pittsburgh

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code): Center for Multivariate Analysis, 515 Thackeray Hall, Pittsburgh PA 15260

7b. ADDRESS (City, State and ZIP Code): Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL (If applicable): NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620-82-K-0001

8c. ADDRESS (City, State and ZIP Code): Bolling AFB DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A5	

11. TITLE (Include Security Classification): BOOTSTRAPPING THE KALMAN FILTER

12. PERSONAL AUTHOR(S): D.S. Stoffer

13a. TYPE OF REPORT: Technical

13b. TIME COVERED: FROM ______ TO ______

14. DATE OF REPORT (Yr., Mo., Day): DEC 84

15. PAGE COUNT: 23

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Bootstrap; Kalman filter; forecasting; robustness.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The bootstrap is proposed as a method for estimating the precision of forecasts and estimates of parameters of the Kalman filter model. It is shown that when the system and the filter is in steady state the bootstrap applied to the Gaussian innovations yields asymptotically consistent standard errors. That the bootstrap works well with moderate sample sizes and supplies robustness against departures from normality is substantiated by emperical evidence.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: MAJ Brian W. Woodruff

22b. TELEPHONE NUMBER (Include Area Code): (202) 767- 5027

22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

END

FILMED

3-85

DTIC